



«Онтологический инжиниринг, Semantic Web и семантические технологии»

Лекция № 2

«Теоретические основы онтологического инжиниринга.

Классификация онтологий»

2013/2014 уч.г.

- лектор доц. Рябова Наталия Владимировна
- лабораторные занятия – асс. Золотухин Олег Викторович
- для студентов специальности 8.05010104, 7.05010104 – Системы искусственного интеллекта
- Харьковский национальный университет радиоэлектроники
- кафедра искусственного интеллекта
- ai@kture.kharkov.ua



Основные вопросы

- Направления в проектировании онтологий
- Принципы классификации онтологий
- Типы онтологий
- Онтологии для обработки ЕЯТ (лексические онтологии)





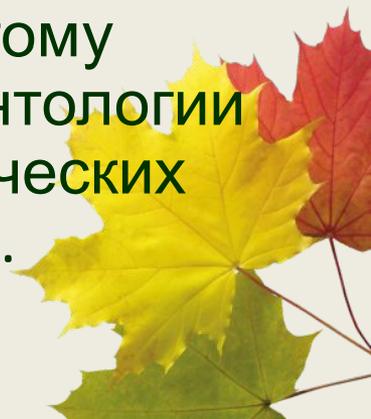
Направления в проектировании онтологий

- 1) представление онтологии как формальной системы, основанной на математически точных аксиомах;
 - 2) развивалось в рамках компьютерной лингвистики и когнитивной науки.
 - Онтология понимается как система абстрактных понятий, существующих только в сознании человека, которая может быть выражена на естественном языке (или средствами какой-то другой системы символов).
 - При этом обычно не делается предположений о точности или непротиворечивости такой системы.
- 



Альтернативные подходы к созданию и исследованию онтологий

- 1) *формальный*, основанный на логике (предикатов первого порядка, дескриптивной, модальной и т. п.).
- 2) *лингвистический*, основанный на изучении естественного языка (в частности, семантики) и построении онтологий на больших текстовых массивах, так называемых корпусах.
- В настоящее время данные подходы тесно взаимодействуют. Идет поиск связей, позволяющих комбинировать соответствующие методы. Поэтому иногда бывает сложно отделить лексические онтологии с элементами формальных аксиоматик от логических систем с включениями лингвистических знаний.





Три основных принципа классификации онтологий

- по степени формальности;
- по наполнению, содержанию;
- по цели создания.





Классификация онтологий по степени формальности

- Обычно люди и программные агенты имеют некоторое представление о значениях терминов.
- Программные агенты иногда предоставляют спецификацию входных и выходных данных, которые также могут быть использованы как спецификация программы.
- Сходным образом онтологии могут быть применены, чтобы предоставить конкретную спецификацию имен терминов и значений терминов.
- В рамках этого понимания (где онтология является спецификацией концептуальной модели — концептуализации) существует простор для вариаций.
- Отдельные виды онтологий могут быть представлены как точки на спектре в зависимости от деталей их реализации



Спектр онтологий. Косая черта разделяет системы, представляющие «машино-понятные»(ниже черты) и «человеко-понятные» (выше черты) описания



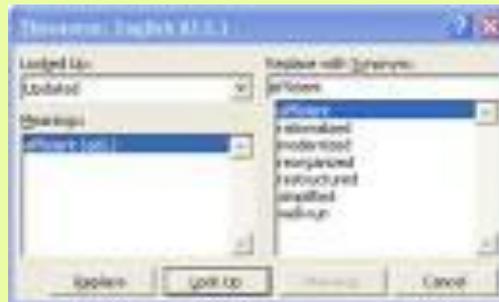
Спектр онтологий. Каталоги на основе ID

- Первой точке на спектре соответствует **контролируемый словарь**, т. е. конечный список терминов (простейшим примером является каталог на основе идентификаторов).
- **Каталоги** представляют точную (не многозначную) интерпретацию терминов.
- Например, каждый раз, ссылаясь на термин "машина", мы будем использовать одно и то же значение (соответствующее некоторому ID в словаре), вне зависимости от того, о чем идет речь в контексте: о "стиральной машине", "автомобиле" или "государственной машине".



Спектр онтологий. Тезаурусы

- **Тезаурусы** несут дополнительную семантику, определяя связи между терминами.
- Отношения, свойственные для тезаурусов:
 - синонимия,
 - иерархическое отношение
 - ассоциация.
- Ранние иерархии терминов, появившиеся в Web, определяли термины через операции обобщения и уточнения.





Спектр онтологий. Тезаурусы

- Yahoo, например, ввела небольшое число категорий верхнего уровня, таких, как "предметы одежды". Затем «платье» определялось как вид (женской) одежды. Явная иерархия Yahoo не соответствовала в точности формальным свойствам иерархического отношения (*isA*).
- В таких иерархиях может встретиться ситуация, в которой экземпляр класса-потомка не является экземпляром класса-предка. Например:
 - общая категория "предметы одежды" имеет подкатегорию "женские" (которая должна была бы более точно называться "женские предметы одежды"),
 - эта категория, в свою очередь, включает подкатегории "аксессуары" и "платье".
 - ясно, что аксессуары, например "броши", не являются предметами одежды. Здесь **не выполняется важное свойство отношения *isA* — транзитивность.**



Спектр онтологий. Формальные таксономии

- Далее следует точка "**формальные таксономии**". Эта разновидность онтологий включает точное определение отношения *isA* ("**КЛАСС-ПОДКЛАСС**").
- В таких системах строго соблюдается *транзитивность* отношения *isA*: если В является подклассом класса А, то каждый подкласс класса В также является подклассом класса А.
- Для отношения "**КЛАСС-ЭКЗЕМПЛЯР**" (*isInstanceOf*) выполняется следующее свойство: если В является подклассом класса А, то каждый экземпляр класса В также является экземпляром класса А.
- Поэтому "**броши**" не могут быть помещены ниже в иерархии "**предмет одежды**", даже в подкатегорию "**женские**", или стать экземпляром этой категории.
- Строгая иерархия необходима при использовании наследования для процедуры логического вывода.





Спектр онтологий.

Наличие формального отношения
«КЛАСС-ЭКЗЕМПЛЯР»

- Некоторые классификации включают только имена классов
- другие содержат на нижнем уровне экземпляры (индивиды).
- Данная точка спектра допускает наличие у классов экземпляров (примеров).





Спектр онтологий.

Свойства на основе фреймов

- Далее среди структурных элементов появляются слоты.
- Здесь классы (иногда их называют фреймами) могут иметь информацию о свойствах (слотах).
- Например, класс "ПРЕДМЕТ ОДЕЖДЫ" может иметь свойства "цена", "сделанИз".
- Свойства бывают особенно полезными, когда они определены на верхних уровнях иерархии и наследуются подклассами.
- В потребительской иерархии класс "ПРОДУКТ" может иметь свойство "цена", которое получают все его подклассы.



Спектр онтологий.

Ограничения на значения

- Большой выразительностью обладают онтологии, включающие *ограничения на область значений свойств*.
- **Значения свойств** берутся из некоторого предопределенного множества (целые числа, символьные константы) или из подмножества концептов онтологии (множество экземпляров данного класса, множество классов).
- Можно ввести дополнительные ограничения на то, что может заполнять свойство.
- Например, для свойства "**сделанИз**" класса «**Предмет одежды**» значения могут быть ограничены экземплярами класса «**Материал**».
- Проблемы, которые могут возникнуть в этом случае при использовании нестрогой таксономии: если "**духи**" — потомок класса «**Предмет одежды**», он унаследует свойство "**сделанИз**" вместе с ограничением ("**Материал**").





Спектр онтологий.

Дизъюнктивные классы. Обратные свойства

- При необходимости описывать более сложные факты выразительные средства онтологии (и ее структура) усложняются.
- Например, может потребоваться заполнить значение какого-либо свойства экземпляра, используя математическое выражение, основанное на значениях других свойств данного экземпляра или значениях свойств других экземпляров.
- Многие онтологии позволяют объявлять два и более класса **дизъюнктивными (непересекающимися)**. Это означает, что у данных классов не существует общих экземпляров.



Спектр онтологий.

Произвольные логические ограничения

- Некоторые языки описания онтологий позволяют делать произвольные логические утверждения о концептах — **аксиомы**.
- Языки описания онтологий, подобные **СусL** и **Ontolingua**, позволяют фиксировать утверждения на языке **логики предикатов первого порядка (FOL)**.





Типы онтологий

- **Онтологии верхнего уровня – Top Level Ontologies;**
- **Онтологии предметных областей - Domain Ontologies;**
Прикладные онтологии - Application Ontologies;
- **Лексические онтологии.**



Классификация онтологий по цели создания (1)

- В рамках этой классификации выделяют четыре уровня:
 - онтологии представления,
 - онтологии верхнего уровня,
 - онтологии предметных областей
 - прикладные онтологии.





Классификация онтологий по цели создания (2)



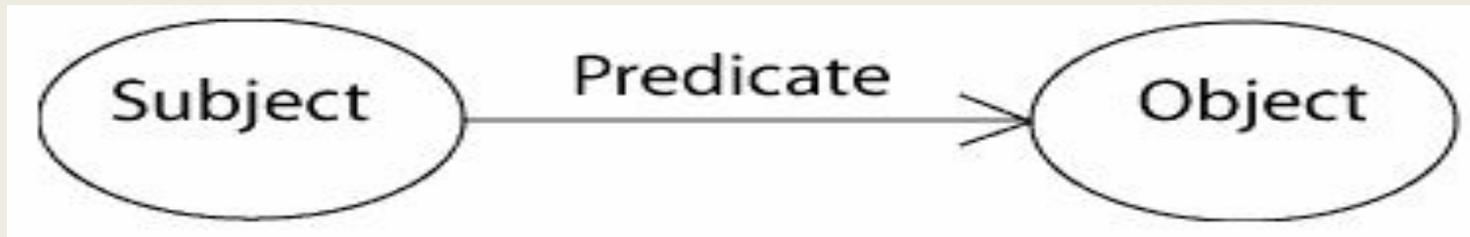


Онтологии представления

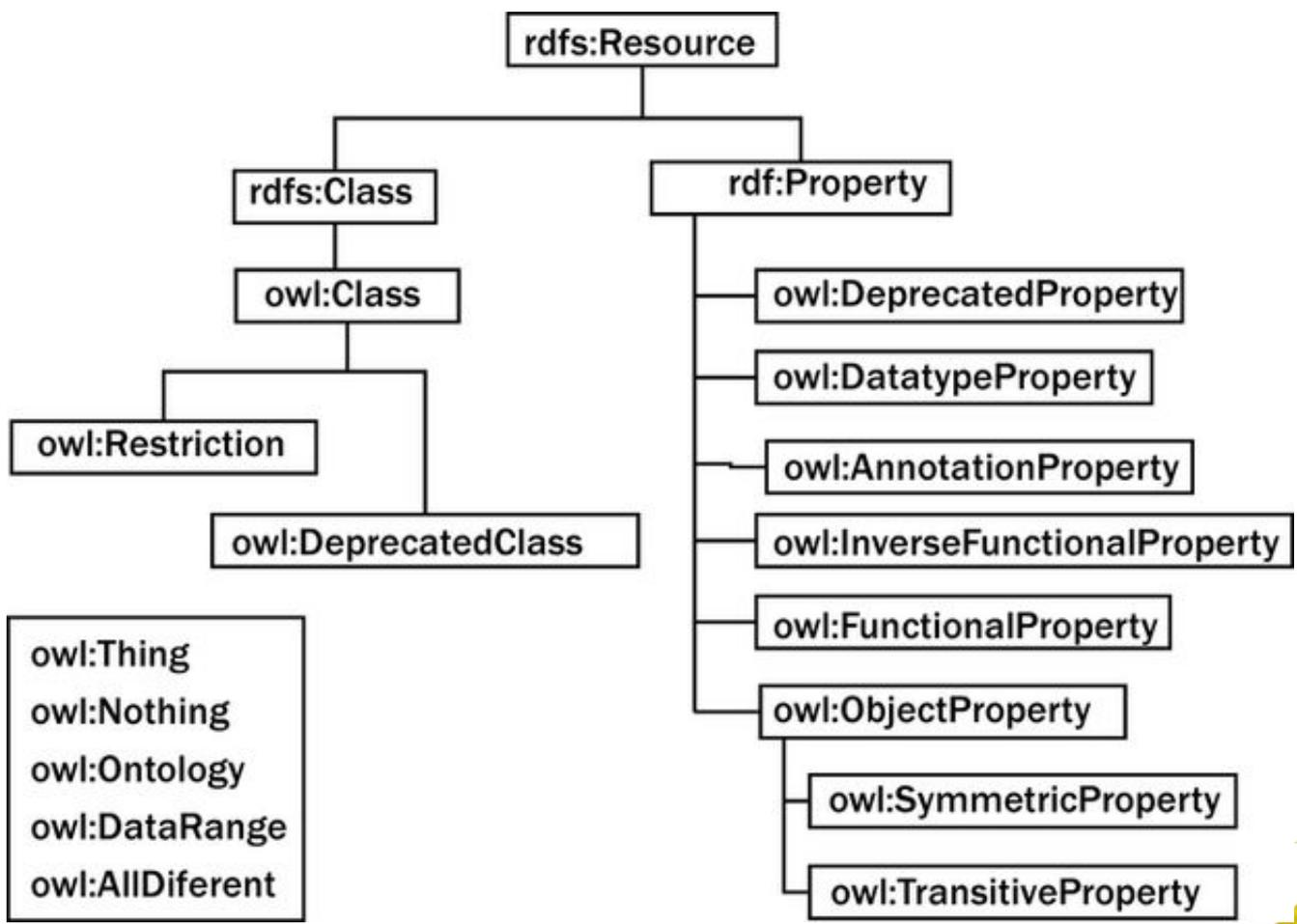
- **Цель их создания:**
 - описать область представления знаний,
 - создать язык для спецификации других онтологий более низких уровней.
 - Пример: описание понятий языка OWL средствами RDF/RDFS.
 - В данном описании определяются такие понятия, как "класс", "отношение", "ограничение на значение свойства", "домен", "диапазон" и т. п.

RDF

Базовой структурной единицей RDF (*Resource Description Framework*) является коллекция троек (или триплетов), каждая из которых состоит из **субъекта, предиката и объекта (S,P,O)**. Набор триплетов называется RDF-графом. В качестве вершин графа выступают субъекты и объекты, в качестве дуг — предикаты (или свойства). Направление дуги, соответствующей предикату в данной тройке (S,P,O), всегда выбирается так, чтобы дуга вела от субъекта к объекту.



Онтология представления для языка OWL

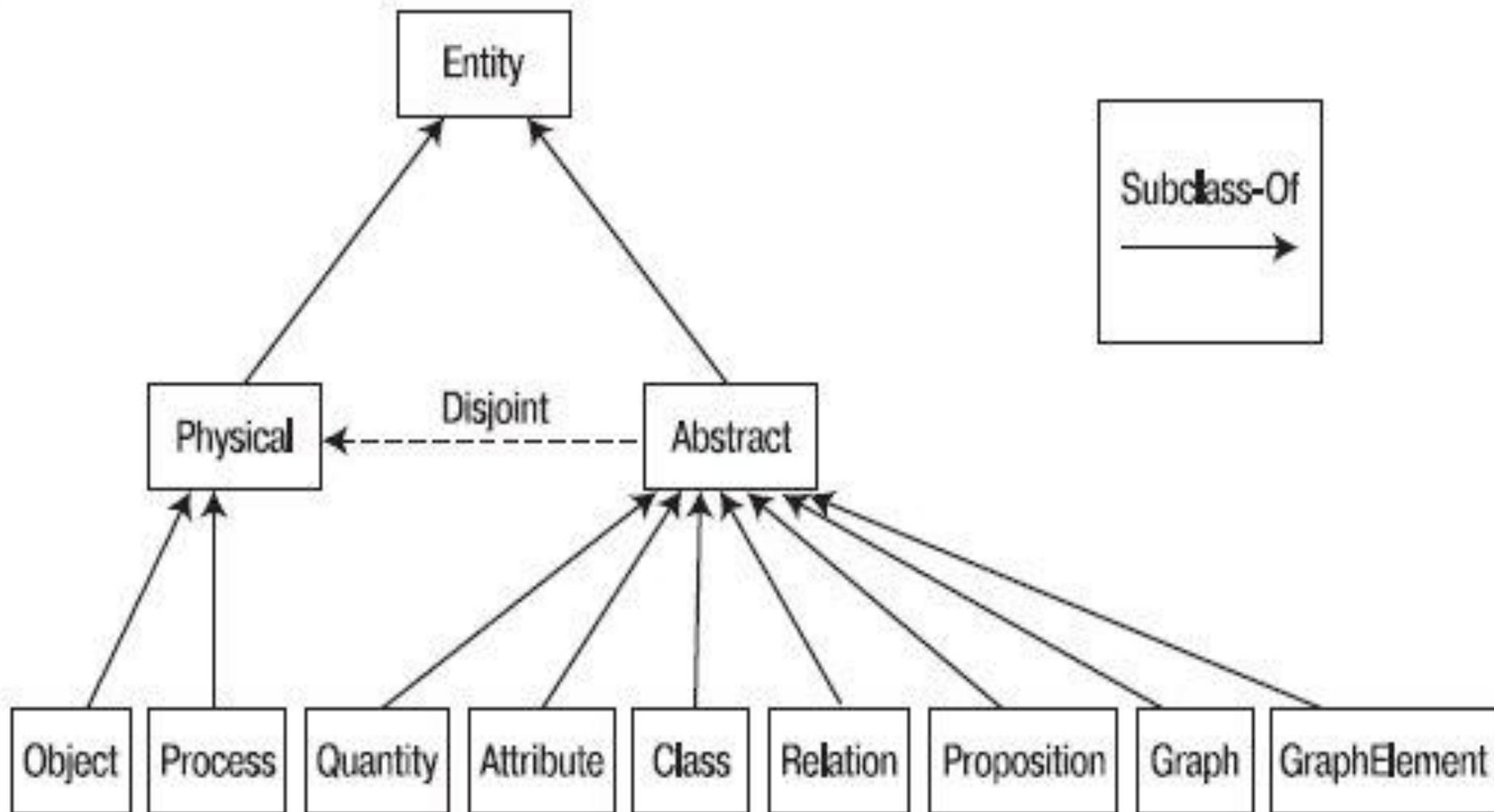




Онтологии верхнего уровня

- Назначение — создание единой "правильной" онтологии, фиксирующей знания, общие для нескольких ПрО, и многократном использовании данной онтологии.
- Существует несколько крупных проектов:
 - SUMO,
 - Sowa's Ontology,
 - Сус,
 - DOLCE и др.
- Попытки создать онтологию верхнего уровня на все случаи жизни пока не привели к ожидаемым результатам.
- Многие онтологии верхнего уровня похожи друг на друга. Они содержат одни и те же концепты:
 - Сущность,
 - Явление,
 - Процесс,
 - Объект,
 - Роль и т. п.

Верхний уровень иерархии онтологии SUMO (Standardized Upper Merged Ontology)



DOLCE – Descriptive Ontology for Linguistic and Cognitive Engineering

- Laboratory for Applied ontology (<http://loa.istc.cnr.it/>)

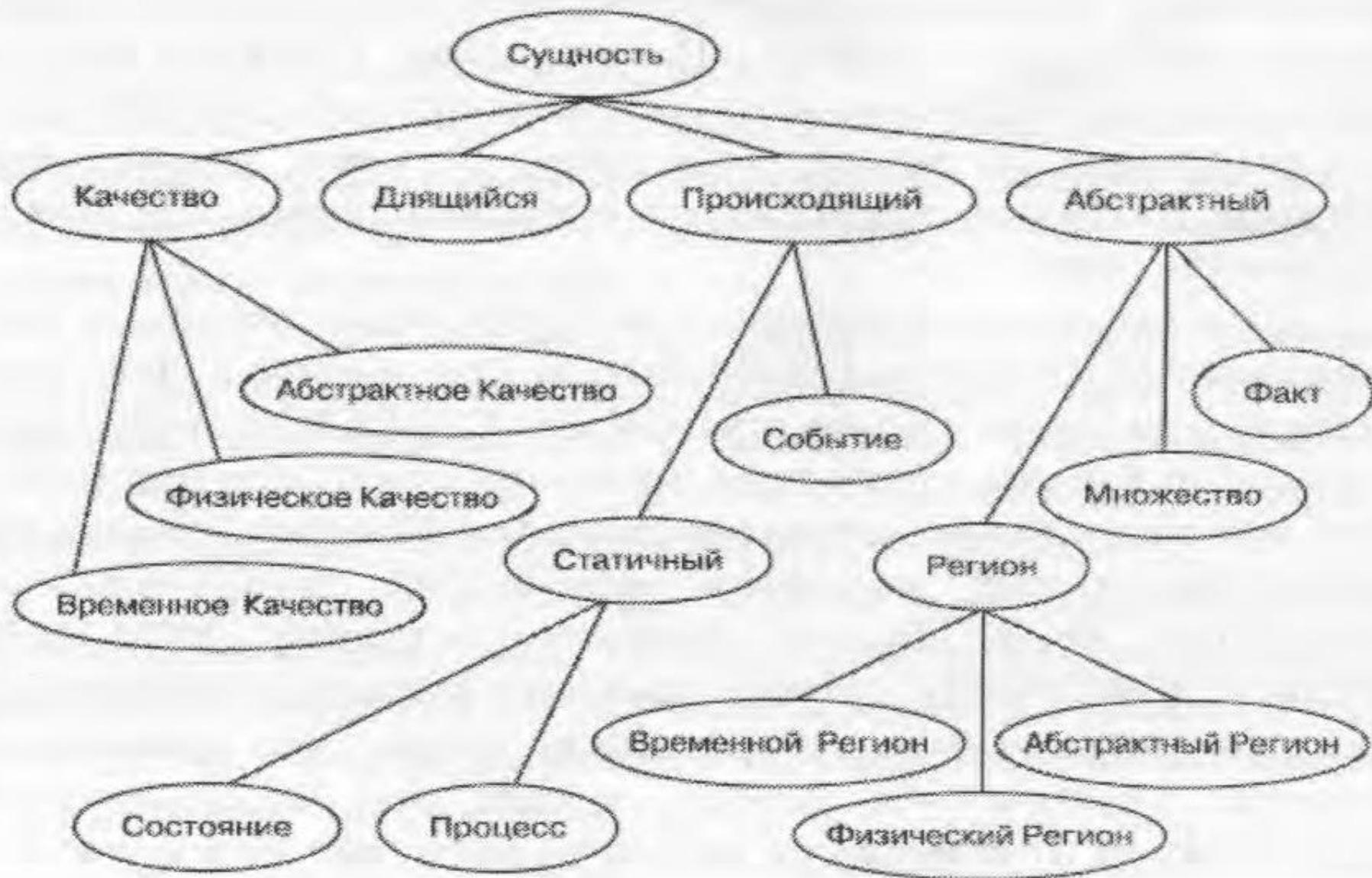
Wonder



Web



Верхние уровни иерархии DOLCE





Онтологии предметных областей (другое название — онтология домена)

- Назначение схоже с назначением онтологий верхнего уровня, но область интереса ограничена предметной областью (т.н. доменом), например, авиация, медицина, культура, дистанционное обучение, Интернет-технологии.
- Онтология предметной области обобщает понятия, использующиеся в некоторых задачах домена, абстрагируясь от самих задач (так, онтология автомобилей независима от любых особенностей конкретных марок машин).
- Примеры:
 - АвиаОнтология,
 - CIDOC CRM,
 - UMLS.



Прикладные онтологии

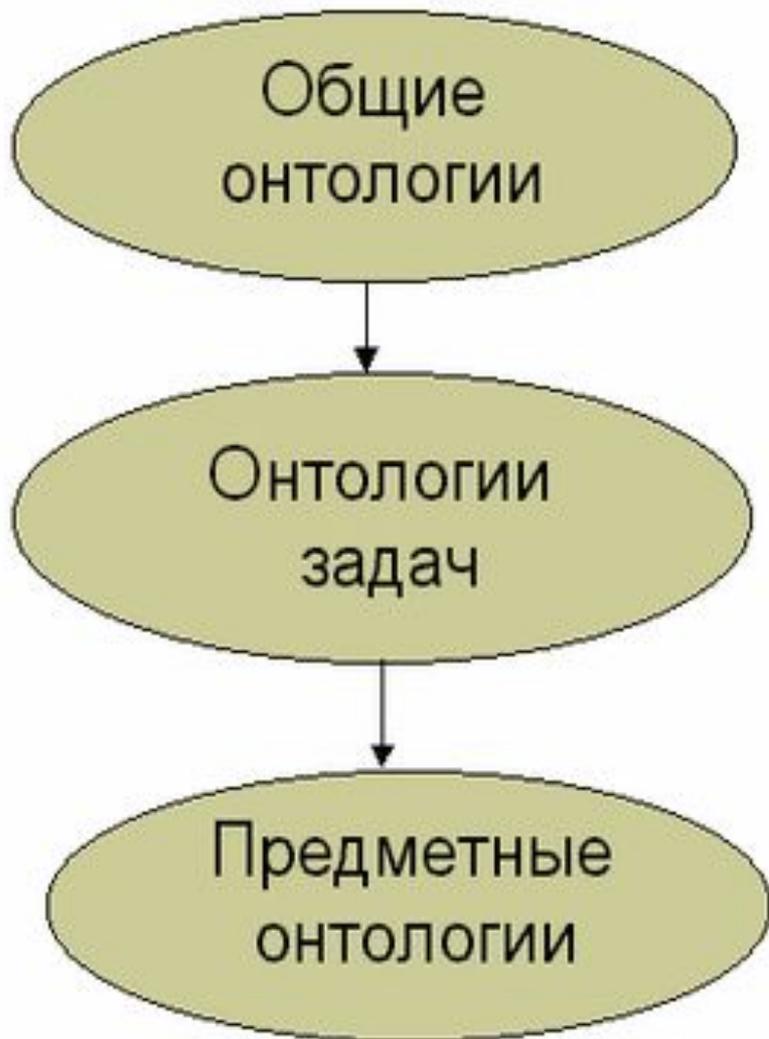
- Назначение этих онтологий в том, чтобы описать *концептуальную модель конкретной задачи или приложения.*
- Прикладные онтологии описывают концепты, которые зависят как от онтологии задач, так и от онтологии предметной области. Примером может служить онтология для автомобилей, строительных материалов, вычислительной техники.
- Такие онтологии содержат наиболее специфичную информацию.

Примеры проектов:

- TOVE,
- Plinius.



Классификация онтологий по содержанию



Сущности, События,
Пространство, Время...

Составление расписаний,
определение целей,
классификация

Множества предметов: скальпели,
сканеры



Классификация онтологий по содержанию

- **Общие онтологии** описывают наиболее общие концепты (пространство, время, материя, объект, событие, действие и т.д.), которые независимы от конкретной проблемы или области.
- **Онтология, ориентированная на задачу** — это онтология, используемая конкретной прикладной программой и содержащая термины, которые используются при разработке ПО, выполняющего конкретную задачу. Задачи могут быть самыми разнообразными: составления расписания, определение целей, диагностика, продажа, разработка ПО, построение классификации.
- **Предметная онтология** (или онтология предметов) описывает реальные предметы, участвующие в какой-либо деятельности (производстве). Например, онтология всех частей и компонентов самолетов определенной марки (Boeing) и сведения об их поставщиках, характеристиках, способе соединения друг с другом и т.п.



Онтологии для обработки текстов на естественном языке.

Лексические онтологии





Актуальность

- Чтобы применить онтологию для автоматической обработки текстов, в частности, для решения задач информационного поиска, необходимо понятиям онтологии сопоставить набор языковых выражений (слов и словосочетаний), которыми понятия могут выразиться в тексте.
- Процедура сопоставления понятий онтологии и языковых выражений может быть осуществлена различными способами.



Процедура сопоставления понятий онтологий и языковых выражений

- Во-первых, онтология может быть сделана заранее, путем логической классификации, а затем к ее элементам могут быть приписаны языковые единицы.
- Так, например, Дуг Ленат (Doug Lenat), руководитель известного проекта в области представления знаний Сус, в рамках которого предполагалось формализовать знания здравого смысла (common sense) и использовать их, в частности, для обработки ЕЯТ, считает, что:
 - учет значений слов может только запутать ("words are often red herrings"),
 - значения слов делят мир неоднозначно, а линии деления происходят из самых различных причин: исторических, физиологических и т. п.



Подход Top-Down

- Предлагается создавать онтологию путем логического анализа, "**сверху-вниз**".
- При этом имена вводимых понятий (желательно) должны отражать те признаки, которые заложены в основу деления.
- В результате получаются имена понятий достаточно громоздкие, неестественные, с ними трудно оперировать как разработчикам, так и возможным пользователям.



Проблемы подхода Top-Down

- Другая проблема такого подхода: приписывая языковые выражения логически обоснованной системе понятий, мы получаем, что одно и то же слово может соответствовать слишком большому количеству таких "правильных" понятий в зависимости от контекста, а значит, возникает излишняя многозначность лексической единицы.
 - Небольшие онтологии могут быть построены методом сверху-вниз, но разработка подробных онтологий для реальных приложений — задача непростая.
- 



- **Более того, во многих ПрО знание, нужное для распространения и интеграции, содержится в основном в текстах.**
- **Из-за внутренних свойств человеческого языка непростой задачей является связать знания, содержащиеся в текстах, с онтологиями, даже если бы была построена подробная онтология ПрО.**





- Некоторые исследователи, такие как известный британский лингвист Йорик Вилкс, считают, что "несмотря на то, что все авторы статей по онтологиям подчеркивают, что понятия являются кирпичиками любой онтологии, мы манипулируем понятиями посредством слов. Во всех онтологиях, которые известны, слова используются для того, чтобы представлять понятия. Следовательно, то множество явлений в мире, которые не вербализованы, не могут быть смоделированы. Мы можем описать это явление как Онтологическую гипотезу Сепира-Уорфа, то есть то, что не описывается словами, не может быть отражено в онтологии...".

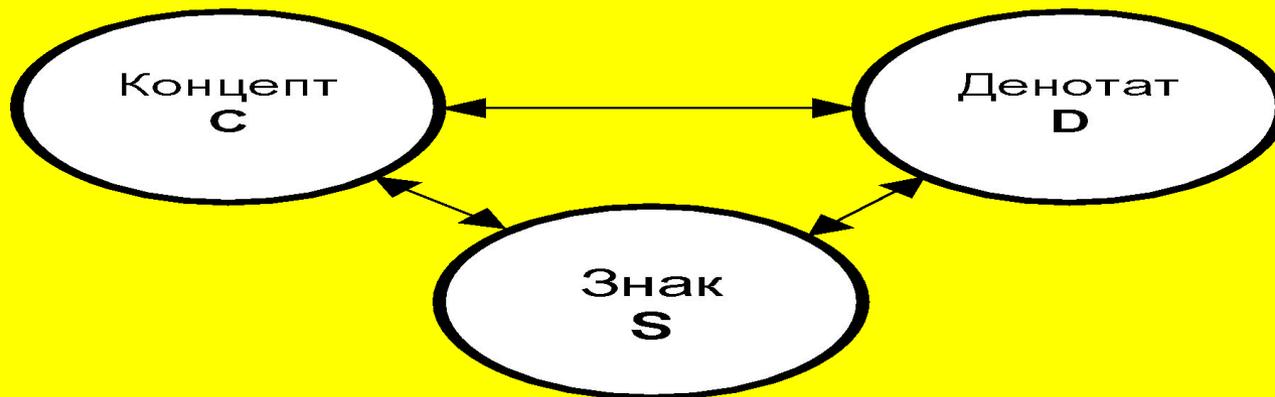


Различие в абстрактных моделях разных субъектов при рассмотрении одного объекта



Д. В. Кудрявцев «Системы управления знаниями и применение онтологий». - Санкт-Петербург, Изд-во Политехнического университета, 2010

Треугольник Фреге

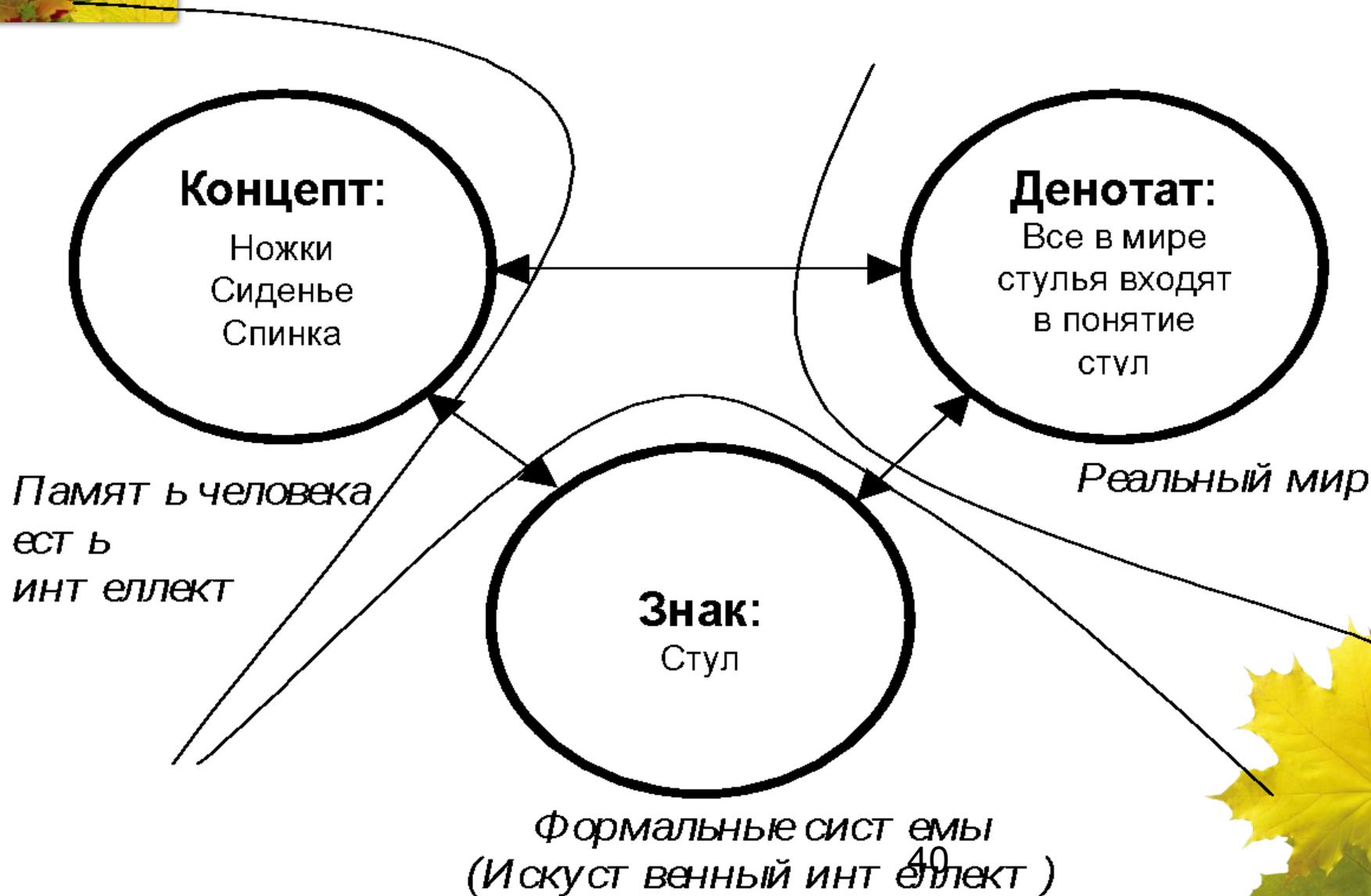


«ЗНАК» – материальный, чувственно воспринимаемый объект, который символически, условно представляет и отсылает к обозначаемому им предмету, явлению, действию или событию.

«КОНЦЕПТ» (понятие) - определение обозначаемого посредством *знака* объекта.

«ДЕНОТАТ» – любой предмет, процесс, явление, составляющее содержание языкового (*знакового*) выражения и определяемого посредством *денотата*. Иными словами, - сущность, явление реального или ментального мира, кодируемая языковым знаком.

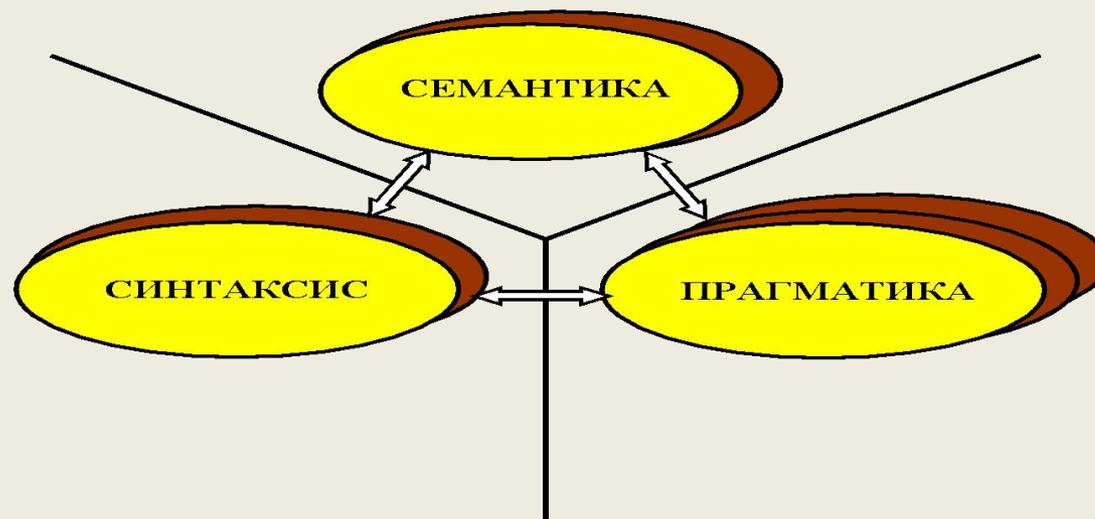
Представление понятия «стул» в знаковой системе



систем,

семантику, рассматривающую смысл (интерпретацию) знаковых систем (другими словами, соответствие знаковой системы другой знаковой системе),

прагматику, имеющую дело с целенаправленностью знаковых систем.



В семиотике различают следующие разделы:

синтактику, имеющую дело со структурой (синтаксисом) знаковых систем,

семантику, рассматривающую смысл (интерпретацию) знаковых систем (другими словами, соответствие знаковой системы другой знаковой системе),

прагматику, имеющую дело с целенаправленностью знаковых систем.

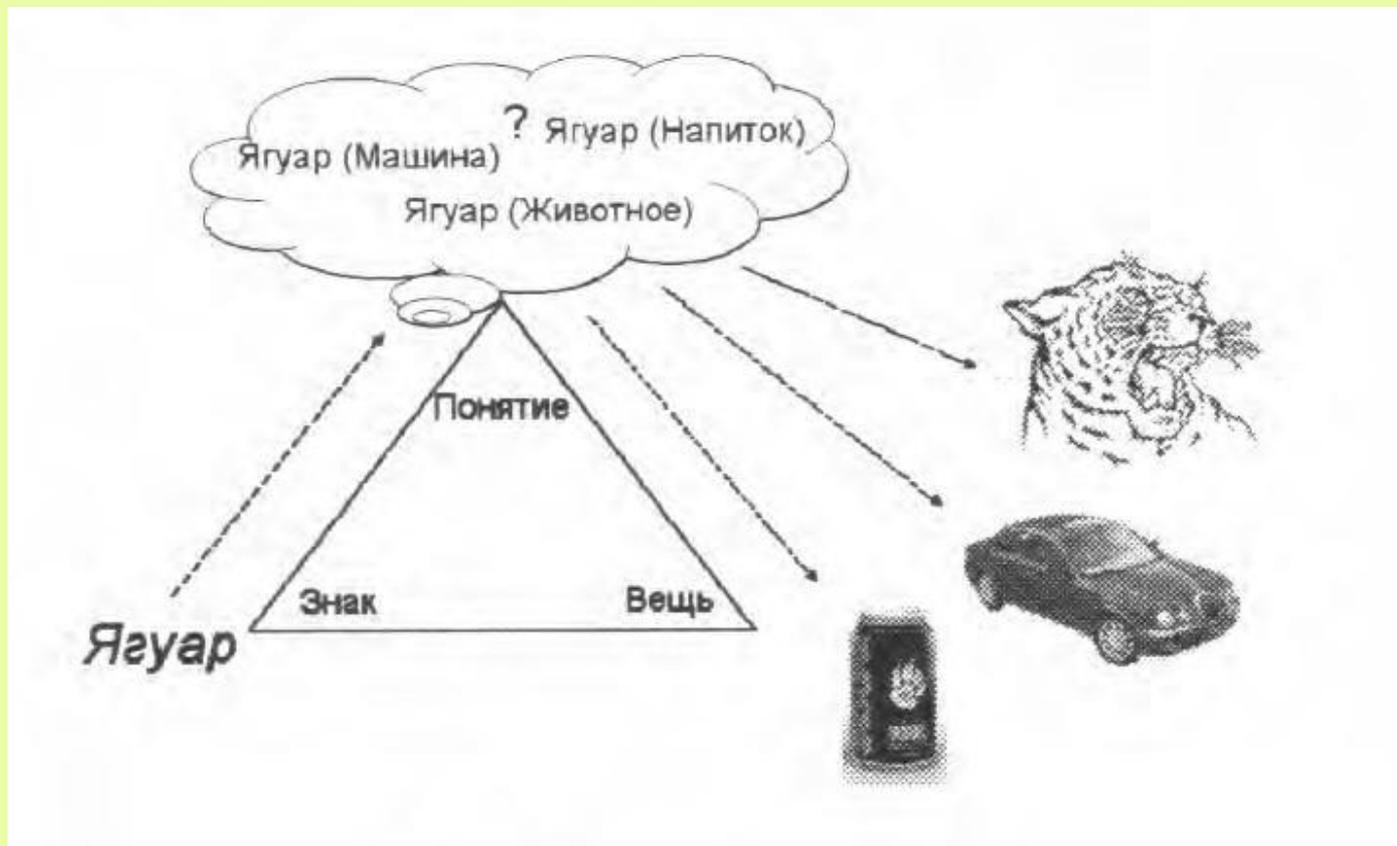
Семантический треугольник



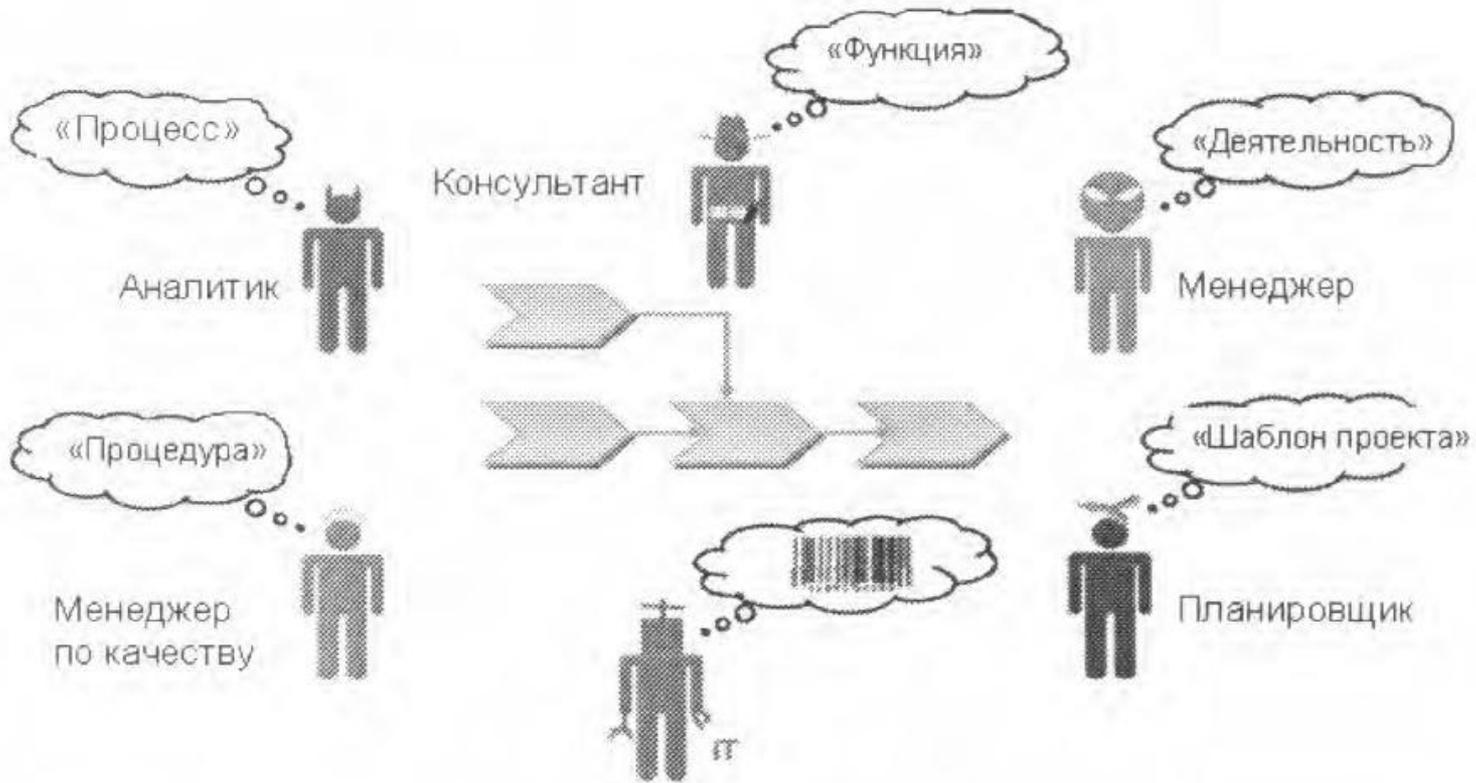
**В этом треугольнике часто
возникают неоднозначности:**

**Синонимия,
Омонимия,
Полисемия.**

Пример омонимии



Пример распространения синонимии в организации





Процедура сопоставления понятий онтологий и языковых выражений

- Второе направление, которое обычно обсуждается, — это установление соответствий между иерархическими лексическими ресурсами типа WordNet и некоторой онтологией.
- WordNet-ресурсы описывают лексические отношения между значениями слов, представленные в виде отдельных единиц в иерархической сети — **синсетов.**
- Отношения между лексическими единицами в значительной мере отражают отношения объектов внешнего мира, поэтому такие ресурсы часто рассматриваются как особый вид онтологий — лексические или лингвистические онтологии.



Основные характеристики лингвистических онтологий

- Главной характеристикой лингвистических онтологий является то, что они привязаны к значениям ("are bound to the semantics") языковых выражений (слов, именных групп и т.п.).
- Лингвистические онтологии охватывают большинство слов языка и одновременно имеют онтологическую структуру, проявляющуюся в отношениях между понятиями. Поэтому **лингвистические онтологии могут рассматриваться как особый вид лексической базы данных и особый тип онтологии.**
- Лингвистические онтологии отличаются от формальных онтологий по степени формализации. Поэтому предполагается, что разработчики такого рода ресурсов разрабатывают иерархию лексических значений естественного языка, а для более строгого описания знаний о мире необходимо сопоставить такие ресурсы с какими-либо формальными онтологиями.



WordNet&SUMO

- Так, содержанием одного из проектов является установление отношений между **WordNet** и **EuroWordNet**, с одной стороны, и формальной онтологией **SUMO** — **Standardized Upper Merged Ontology** — с другой.
- Проект состоит в том, чтобы установить соответствие между синсетами **WordNet** и понятиями онтологии, при котором каждый синсет **WordNet** либо напрямую сопоставляется с понятием онтологии, либо является гипонимом для некоторого понятия или примером понятия онтологии.





Проект OntoWordNet

- Участники другого проекта — **OntoWordNet** — считают, что недостаточно провести формальную склейку ресурса типа **WordNet** и формальной онтологии: необходима значительная реструктуризация исходного лексического ресурса.
 - Аналогичная работа проводится и в проекте **Сус**.



Процедура сопоставления понятий онтологий и языковых выражений

- *Третий путь* — попытаться разработать единый ресурс, в котором были бы сбалансированы обе части:
 - **система понятий** и
 - **система лексических значений**,
- что заключается в разумном разделении этих единиц в создаваемом ресурсе и аккуратном описании их взаимосвязей.
- Попытка такого подхода реализуется в онтологиях **MikroKosmos** и **OntoSem**.



Контрольные вопросы

- Чем отличаются онтологии верхнего уровня от онтологий предметной области?
- Чем отличаются онтологии предметной области от прикладных онтологий?
- Перечислите основные характеристики лексических онтологий



Thank You!

